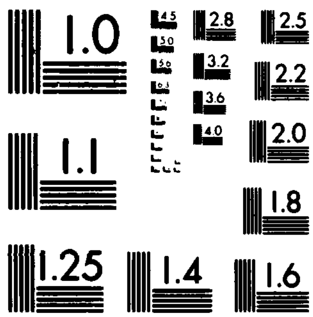
AD-A082 396 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 12/1
KTH ORDER SOLUTIONS TO THE PROBLEM OF FINDING OPTIMAL DISCRIMIN--ETC(U)
JAN 80 L K JONES F19628-80-C-0002
UNCLASSIFIED TN-1980-4 ESD-TR-79-305 NL

END
DATE FILMED
4 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

6

k^{th} ORDER SOLUTIONS TO THE PROBLEM OF FINDING OPTIMAL DISCRIMINANT FUNCTIONS.

10 Lee K. JONES

Group 92

14 TN-1980-4

9 TECHNICAL NOTE 1980-4

11 10 JAN[redacted]980

12 25

A

15 F19628-80-C-0002

16 627A

18 ESD

19 TR-79-305

LEXINGTON MASSACHUSETTS

Abstract

This article describes a general method for discriminating between two random variables when we are required to use a discriminant function belonging to a given class.

Contents

I. Introduction

In this note we consider the "two class" problem of statistical classification: We are given two random variables, X_1 and X_2, taking values in R^d together with some (usually incomplete) information about their distributions. We assume occurrences of type $1(X_1)$ and of type $2(X_2)$ are mutually exclusive and have prior probabilities of α and $1-\alpha$ respectively $(0 < \alpha < 1)$. If x is observed how do we decide if x is of type 1 or of type 2 in such a fashion as to minimize the probability of making an incorrect decision?

If the probability densities of X_1 and X_2, $p_1(y)$ and $p_2(y)$, were known we would decide by using the likelihood ratio test:

$$\frac{\alpha p_2(x)}{(1-\alpha)p_1(x)} \quad \begin{array}{lll} > & 1 & \text{type } 2 \\ \leq & 1 & \text{type } 1 \end{array}$$

Unfortunately in many practical situations (good estimates of) the probability densities are unavailable. However (good estimates of) other statistics are available (lower order moments, spectral estimates, features, etc.). These enable one to construct a family of discriminant functions L whose errors may be (estimated) calculated from the (estimated) known statistics. We give a formal definition of a discriminant function and its error as follows.

Def. 1 A discriminant function L is a mapping $L: R^d \to R$. The error of L (relative to the above classification problem) is the infimum over all real t of the expected probability of error of the decision rule -

$$\begin{array}{ll} L(z) > t & \text{type } 2 \\ L(z) \leq t & \text{type } 1 \end{array}$$

The error of L is given by the expression

$$\inf_{-\infty < t < +\infty} \left[\alpha \, \mathrm{Prob}_1 \left\{ L(X_1) > t \right\} + (1-\alpha) \, \mathrm{Prob}_2 \left\{ L(X_2) \leq t \right\} \right]^*$$

Given a class of discriminants $\mathcal{L}$ our goal is to find an $L \in \mathcal{L}$ of minimum error. In II we will assume that the distributions of $L \in \mathcal{L}$ under each hypothesis may be parametrized by k parameters. Necessary conditions for minimum error will be derived. In III certain properties of the normal distribution relative to the framework of II will be discussed. Finally a third order solution for the optimal linear discriminant will be given in IV.

II. Necessary Conditions for k^{th} Order Solutions

Let $\mathcal{D}$ be a class of continuous probability densities on the real line parametrized by their means, variances, third moments,..,k^{th} moments about the mean $(v^1, v^2, v^3, \ldots, v^k)$. We assume further that $\mathcal{D}$ is a location family: $\mu(v^1 - u^1, v^2, v^3, \ldots, v^k)(x) = \mu(v^1, v^2, v^3, \ldots, v^k)(x + u^1)$. Consider any two densities $\mu(0, v_1^2, v_1^3, \ldots, v_1^k)$ and $\mu_2 = \mu(1, v_2^2, \ldots, v_2^k)$. Let $E_\alpha(\mu_1, \mu_2) = E_\alpha(\mu(0, v_1^2, \ldots), \mu(1, v_2^2, \ldots))$ be the error of the identity discriminant function in R (for the two class problem with densities μ_1, μ_2).

*See Footnote (1) on page 18.

$$E_\alpha(\mu_1, \mu_2) = \inf_t \left\{(1-\alpha)\int_{-\infty}^{t} \mu_2 \, dx + \alpha \int_{t}^{+\infty} \mu_1 \, dx\right\}$$

Assume further that $E_\alpha(\mu_1, \mu_2) = E_\alpha(v_1^{\,2}, v_1^{\,3}, \ldots, v_1^{\,k}; v_2^{\,2}, v_2^{\,3}, \ldots, v_2^{\,k})$ has continuous partials wrt $v_1^{\,2}, v_1^{\,3}, \ldots, v_2^{\,k}$.

<u>Def. 2</u> $\mathcal{D}$ is said to have <u>monotone error</u> at the pair (μ_1, μ_2) if $E_\alpha(v_1^{\,2}, v_1^{\,3}, \ldots, v_2^{\,k})$ has a non-vanishing gradient at $v_1^{\,2}, v_1^{\,3}, \ldots, v_2^{\,k}$.

Let $A \subset R^q$ be some parameter space and let $L = \{L(\vec{a}); \vec{a} \in A\}$ be a family of discriminant functions with

$$E_2(L(\vec{a})) - E_1(L(\vec{a})) = 1 \text{ for all } \vec{a}.$$

Suppose the probability density of $L(\vec{a})$ under each hypothesis lies in $\mathcal{D}$ and the mapping $\vec{A} \to E_\alpha(v_1^{\,2}(\vec{a}), \ldots\, v_1^{\,k}(\vec{a}); v_2^{\,2}(\vec{a}), \ldots, v_2^{\,1}(\vec{a}))$ has partial derivatives of the first order. ($v_i^{\,j}(\vec{a})$ is the k^{th} moment about the mean of the random variable $L(\vec{a})$ under hypothesis i.)

<u>Theorem 1</u> Let $\mathcal{D}$ have monotone error at $(\mu(0, v_1^{\,2}(\vec{a}'), \ldots\, v_1^{\,k}(\vec{a}')), \mu(1, v_2^{\,2}(\vec{a}'), \ldots, v_2^{\,k}(\vec{a}'))$. If $E_\alpha(v_1^{\,2}(\vec{a}), \ldots, v_1^{\,k}(\vec{a}); v_2^{\,2}(\vec{a}), \ldots, v_2^{\,k}(\vec{a}))$ has a local minimum at $\vec{a}'$, then $\vec{a}'$ is a critical point of $\sum_{i=1}^{2} \sum_{j=2}^{k} \beta_i^{\,j} v_i^{\,j}(\vec{a})$ for some set of $2(k-1)$ real numbers $\beta_i^{\,j}$ (not all 0) with $-1 \leq \beta_i^{\,j} \leq +1$. If $E_\alpha(v_1^{\,2}, \ldots\, v_2^{\,k})$ is strictly concave (as a function of $v_1^{\,2}, \ldots\, v_2^{\,k}$) at $v_1^{\,2}(\vec{a}'), \ldots, v_2^{\,k}(a')$, then the above critical point $\vec{a}'$ is a strict local minimum of $\sum_{i=1}^{2} \sum_{j=2}^{k} \beta_i^{\,j} v_i^{\,j}(\vec{a})$.

Proof Taking partial derivatives of E_α wrt a_r at $\vec{a}'$ we have

$$\sum_{i=1}^{2} \sum_{j=2}^{k} \left(\left.\frac{\partial E_\alpha}{\partial v_i^j}\right|_{\vec{a}'} \right) \left(\left.\frac{\partial v_i^j}{\partial a_r}\right|_{\vec{a}'} \right) = 0$$

Since the gradient of E_α wrt $v_1^2, \ldots v_2^k$ is non-zero at $v_1^2(\vec{a}'), \ldots$ we may set

$$\beta_i^j = \left(\left.\frac{\partial E_\alpha}{\partial v_i^j}\right|_{\vec{a}'} \right) \left(\sum_{i=1}^{2} \sum_{j=2}^{k} \left| \left.\frac{\partial E_\alpha}{\partial v_i^j}\right|_{\vec{a}'} \right| \right)^{-1}$$

Then $\vec{a}'$ is a critical point of $\sum_{i=1}^{2} \sum_{j=2}^{k} \beta_i^j v_i^j(\vec{a})$.

Suppose $E_\alpha(v_1^2, \ldots, v_2^k)$ is strictly concave at $v_1^2(\vec{a}'), \ldots, v_2^k(\vec{a}')$. Since the partial derivatives of E_α wrt v_i^j are continuous, $E_\alpha(v_1^2, \ldots, v_2^k)$ has a differential at $v_1^2(\vec{a}'), \ldots, v_2^k(\vec{a}')$. Hence for any direction $\hat{u}$ ($\|\hat{u}\| = 1$), $\left.\frac{\partial E_\alpha}{\partial \hat{u}}\right|_{\vec{a}'} = \left(\text{grad } E_\alpha \Big|_{\vec{a}'} \right) \bullet \hat{u}$. By strict concavity $E_\alpha(v_i^j(\vec{a}') + \rho u_i^j) - E_\alpha(v_i^j(\vec{a}')) < \rho(\text{grad } E_\alpha \big|_{\vec{a}'}) \bullet \hat{u}$ for ρ sufficiently small, but positive. Hence for $\vec{a}$ sufficiently close to $\vec{a}'$, but unequal $\vec{a}'$,

$$0 \leq E_\alpha(v_i^j(\vec{a})) - E_\alpha(v_i^j(\vec{a}')) = E_\alpha(v_i^j(\vec{a}') + (v_i^j(\vec{a}) - v_i^j(\vec{a}')))$$
$$- E_\alpha(v_i^j(\vec{a}')) < \sum_{i=1}^{2} \sum_{j=2}^{k} \left(\left.\frac{\partial E_\alpha}{\partial v_i^j}\right|_{\vec{a}'} \right) \left(v_i^j(\vec{a}) - v_i^j(\vec{a}') \right)$$

Hence $\vec{a}'$ is a strict local minimum of $\sum_{i=1}^{2} \sum_{j=2}^{k} \beta_i^j v_i^j(\vec{a})$.

The preceding theorem allows us to reduce the parameters in our problem from q to 2(k - 1) as follows: For any choice of reals β_i^j, $-1 \leq \beta_i^j \leq 1$, we find a set of critical points $A(\beta_i^j)$ of $\sum_{i=1}^{2} \sum_{j=2}^{k} \beta_i^j \, v_i^j(\vec{a})$. Then the $L \in \mathcal{L}$ of minimum error is in the set

$$\left\{L(\vec{a}) \; : \; \vec{a} \in A(\beta_i^j), \; -1 \leq \beta_i^j \leq 1\right\}.$$

For each such L in the above set we (estimate) calculate the error from (performance on sample data) knowledge of the densities in $\mathcal{D}$. The L of minimum error is then found by a numerical search in the 2(k - 1) dimensional set described by the β_i^j. Knowledge that the above critical points are indeed strict local minima may be extremely useful for numerical purposes since the number of critical points of $\sum_{i=1}^{2} \sum_{j=2}^{k} \beta_i^j \, v_i^j(\vec{a})$ may be prohibitively large but the number of strict local minima computationally feasible. This will be the case in IV. Hence the concavity condition may be extremely important. For this reason we discuss the strict concavity of $E_\alpha(v_1^{\,1}, v_2^{\,2})$ for k = 2 and $\mathcal{D}$ the set of normal distributions in III.

For the case k = 2 (second order solution) we may reduce our problem to one with a single parameter: determine critical points of $\beta v_1^{\,2}(\vec{a}) + (1 - |\beta|) v_2^{\,2}(\vec{a})$ for $-1 \leq \beta \leq +1$. This was shown in (1) and (2) and applied to various classes of discriminant functions.

The above results are completely analogous when we parametrize $\mathcal{D}$ by statistics other than moments. The choice of such statistics will influence considerably the performance of a k^{th} order solution.

III. Some Remarks on the Normal Distribution

Let $\mathcal{D}$ be the 2-parameter class of normals. For convenience denote the two variance parameters, v_1^2 and v_2^2, by w and z respectively. Let

$$E_\alpha^{\,c}(w, z) = \int_c^{+\infty} \frac{\alpha}{\sqrt{2\pi w}} \exp\left(\frac{-x^2}{2w}\right) dx + \int_{-\infty}^{c} \frac{1-\alpha}{\sqrt{2\pi z}} \exp\left(\frac{-(x-1)^2}{2z}\right) dx$$

Then

$$E_\alpha(w, z) = \inf_c E_\alpha^{\,c}(w, z) = E_\alpha^{\,c(w, z)}(w, z)$$

where

$$c(w, z) = \frac{\frac{2}{z} - \sqrt{\frac{4}{z^2} - 4\left(\frac{1}{z} - \frac{1}{w}\right)\left(\log \frac{\alpha^2 z}{(1-\alpha)^2 w} + \frac{1}{z}\right)}}{2\left(\frac{1}{z} - \frac{1}{w}\right)} \qquad \text{for } z \neq w$$

$$c(w, z) = \frac{z}{2} \log \frac{\alpha^2}{(1-\alpha)^2} + \frac{1}{2} \qquad \text{for } z = w$$

The function c(w, z) has a Taylor expansion about any point of the form (w, w). $(w > 0)$ Hence c(w, z) has continuous partial derivatives of the first order in z and w. It represents the smaller root of the equation $\alpha\mu(0,w) = (1-\alpha)\mu(1,z)$ for $z < w$, the larger root for $z > w$, and the only root for $z = w$.

<u>Lemma 1</u> $E_\alpha(w, z)$ has continuous partial derivatives of the first order given by the formulae

$$\frac{\partial E_\alpha}{\partial w} = \int_{c(w,z)}^{+\infty} \frac{\partial}{\partial w} \left[\frac{\alpha}{\sqrt{2\pi w}} \exp\left(\frac{-x^2}{2w}\right)\right] dx$$

$$\frac{\partial E_\alpha}{\partial z} = \int_{-\infty}^{c(w,\ z)} \frac{\partial}{\partial z} \left[\frac{(1-\alpha)}{\sqrt{2\pi z}} \ \exp\left(\frac{-(x-1)^2}{2z}\right) \right] dx$$

<u>Proof:</u> From the formulae for $\frac{\partial F_\alpha}{\partial w}$ and $\frac{\partial E_\alpha}{\partial z}$ it follows that the derivatives are continuous in w, z. Hence we need only derive the formulae. We derive the expression for $\frac{\partial E_\alpha}{\partial w}$ · $\frac{\partial E_\alpha}{\partial z}$ is derived analogously. Consider

$$\frac{E_\alpha^{\ c(w',\ z)} - E_\alpha^{\ c(w,\ z)}}{w' - w}$$

$$= \frac{1}{w' - w} \int_{c(w,\ z)}^{+\infty} \left[\frac{\alpha}{\sqrt{2\pi w'}} \ \exp\left(\frac{-x^2}{2w'}\right) - \frac{\alpha}{\sqrt{2\pi w}} \ \exp\left(\frac{-x^2}{2w}\right) \right] dx$$

$$+ \frac{1}{w' - w} \int_{c(w',\ z)}^{c(w,\ z)} \left[\frac{\alpha}{\sqrt{2\pi w'}} \ \exp\left(\frac{-x^2}{2w'}\right) - \frac{(1-\alpha)}{\sqrt{2\pi z}} \ \exp\left(\frac{-(x-1)^2}{2z}\right) \right] dx$$

For $\varepsilon > 0$ and w' sufficiently close to w, the integrand in the second term of the previous expression will be of magnitude less than ε. Hence the second term is bounded in absolute value by

$\frac{|c(w',\ z) - c(w,\ z)|}{|w' - w|} \ \varepsilon$ which converges to $\varepsilon \left|\frac{\partial c}{\partial w}\right|$ as $w' \to w$. Since ε was arbitrary the second term converges to zero and the first term converges to the desired expression.

<u>Lemma 2</u> $\mathcal{D}$ has monotone error at each pair (w, z)

<u>Proof</u>: for fixed w, z

$$\frac{\partial E_\alpha}{\partial w} = A \int_{c(w,\ z)}^{+\infty} \left(\frac{x^2}{w} e^{-\frac{x^2}{2w}} - e^{-\frac{x^2}{2w}}\right) dx$$

$$\frac{\partial E_\alpha}{\partial z} = B \int_{-\infty}^{c(w,\ z)} \left(\frac{(x-1)^2}{z} e^{-\frac{(x-1)^2}{2z}} - e^{-\frac{(x-1)^2}{2z}}\right) dx$$

where A and B are non-zero. The first integral vanishes only if $c(w, z) = 0$ and the second only if $c(w, z) = 1$. Hence both partial derivatives are not simultaneously zero.

<u>Theorem 2</u> $E_\alpha(w, z)$ is strictly concave in the region described by the inequalities

$$0 < c(w, z) < 1$$

$$w > \frac{1}{3}\left[c(w, z)\right]^2$$

$$z > \frac{1}{3}\left[1-c(w, z)\right]^2$$

Proof: let $(\hat{w}, \hat{z})$ lie in the above open region. We will show that E_α is strictly concave in a neighborhood of $(\hat{w}, \hat{z})$. There are neighborhoods N of $c(\hat{w}, \hat{z})$ and Δ of $(\hat{w}, \hat{z})$ such that for any $c \in N$ and $(w, z) \in \Delta$

$$0 < c < 1$$

$$w > \frac{1}{3} c^2$$

$$z > \frac{1}{3} (1-c)^2$$

Choose a neighborhood of $(\hat{w}, \hat{z})$, $\Delta^1 \subset \Delta$, such that $c(w, z) \in N$ for all $(w, z) \in \Delta^1$.

In Δ^1

$$E_\alpha(w, z) = \inf_c E_\alpha^{\,c} (w, z) = \inf_{c \in N} E_\alpha^{\,c} (w, z) = E_\alpha^{\,c(w, z)}(w, z).$$

Now it may be easily shown that the infimum of a collection of strictly concave functions, defined in a common open domain, is strictly concave in that domain provided the infimum is assumed at each point in the domain by some element in the collection. Hence we need only show that $E_\alpha^{\,c}(w, z)$ is strictly concave in Δ^1 for all $c \in N$.

We have

$$E_\alpha^{\,c}(w, z) = \frac{\alpha}{\sqrt{2\pi}} \int_{\frac{c}{\sqrt{w}}}^{+\infty} e^{-\frac{x^2}{2}} dx + (1-\alpha) \left[1 - \frac{1}{\sqrt{2\pi}} \int_{\frac{c-1}{\sqrt{z}}}^{+\infty} e^{-\frac{x^2}{2}} dx \right]$$

$$\frac{\partial E_\alpha^{\,c}}{\partial w} = \frac{\alpha c w^{-\frac{3}{2}}}{2\sqrt{2\pi}} e^{-\frac{1}{2}\frac{c^2}{w}}$$

$$\frac{\partial^2 E_\alpha^{\,c}}{\partial w^2} = \frac{c\alpha w^{-\frac{7}{2}}}{4\sqrt{2\pi}} e^{-\frac{1}{2}\frac{c^2}{w}} (c^2 - 3w)$$

$$\frac{\partial E_\alpha{}^c}{\partial z} = \frac{(1-\alpha)(1-c)z^{-\frac{3}{2}}}{2\sqrt{2\pi}} \; e^{-\frac{1}{2}\frac{(1-c)^2}{z}}$$

$$\frac{\partial^2 E_\alpha{}^c}{\partial z^2} = \frac{(1-c)(1-\alpha)z^{-\frac{7}{2}}}{4\sqrt{2\pi}} \; e^{-\frac{1}{2}\frac{(1-c)^2}{z}} \; ((1-c)^2 - 3z)$$

$$\frac{\partial E_\alpha{}^c}{\partial w \partial z} = 0$$

Since the second partials wrt w and z are negative in Δ^1, $E_\alpha{}^c$ is strictly concave.

By inspection of the second partials in the preceding proofs one determines immediately two other regions of strict concavity:

$$\begin{array}{lcl} c < 0 & & c > 1 \\ w < \frac{1}{3}c^2 & \text{and} & w > \frac{1}{3}c^2 \\ z > \frac{1}{3}(1-c)^2 & & z < \frac{1}{3}(1-c)^2 \end{array}$$

<u>Corollary 1</u> Let $L = \{L(\vec{a});\ \vec{a} \in A\}$ be a family of discriminant functions with the properties of section II whose densities are the two parameter family of normals. For convenience denote $v_1{}^2(\vec{a})$, $v_2{}^2(\vec{a})$ by $w(\vec{a})$, $z(\vec{a})$ respectively. If $\vec{a}'$ is a local minimum of $E_\alpha(w(\vec{a}), z(\vec{a}))$ satisfying

$0 < c(w(\vec{a}'), z(\vec{a}')) < 1$

$w(\vec{a}') > \frac{1}{3}\left[c(w(\vec{a}'), z(\vec{a}')\right]^2$

$z(\vec{a}\ ) > \frac{1}{3}\left[1 - c(w(\vec{a}'), z(\vec{a}'))\right]^2,$

then $\vec{a}'$ is a strict local minimum of $\beta w(\vec{a}) + (1-\beta)z\ (\vec{a})$ for some $0 < \beta < 1$.

<u>Proof</u> Since $E_\alpha^{\ 1}$ is strictly concave at $(w(\vec{a}'), z(\vec{a}'))$ by Theorem 2, $\vec{a}'$ is a strict local minimum of some weighted sum of $w(\vec{a})$ and $z(\vec{a})$, $\beta_1^{\ 2}\ w(\vec{a}) + \beta_2^{\ 2}\ z(\vec{a})$. Let $\hat{c} = c(w(\vec{a}'), z(\vec{a}'))$. From Lemma 1

$$\left.\frac{\partial E_\alpha}{\partial w}\right|_{\vec{a}'} = \left.\frac{\partial}{\partial w}\,(E_\alpha^{\ \hat{c}})\right|_{\vec{a}'}$$

$$\left.\frac{\partial E_\alpha}{\partial z}\right|_{\vec{a}'} = \left.\frac{\partial}{\partial z}\,(E_\alpha^{\ \hat{c}})\right|_{\vec{a}'}$$

From the formulae in the proof of Theorem 2 these partial derivatives are both positive. Hence $\beta_1^{\ 2} = \beta$ and $\beta_2^{\ 2} = 1-\beta$ for some $0 < \beta < 1$ from the formulae for $\beta_i^{\ j}$ in the proof of Theorem 1.

<u>Corollary 2</u> Let L be as in Corollary 1. Let $\vec{a}'$ be a local minimum of $E_\alpha(w(\vec{a}), z(\vec{a}))$. Let ε_i be the probability of error of type i for $L' = L(\vec{a}')$. ($\varepsilon_1 = \text{Prob}_1(L' \geq \hat{c})$, $\varepsilon_2 = \text{Prob}_2(L' < \hat{c})$.) Then if ε_i satisfies the inequalities

$$.5 > \varepsilon_i > \int_{\sqrt{3}}^{\infty} \frac{1}{\sqrt{2\pi}}\ e^{-\frac{x^2}{2}}\ dx\ (\approx .04),$$

$\vec{a}'$ is a strict local minimum of $\beta\, w(\vec{a}) + (1-\beta)\, z(\vec{a})$ for some $0 < \beta < 1$.

Proof Since the error probabilities of each type are less than ½, $0 < \hat{c} < 1$. Also

$$\varepsilon_1 = \int_{\hat{c}}^{\infty} \frac{1}{\sqrt{2\pi\, w(\vec{a}')}}\, e^{-\frac{x^2}{2w(\vec{a}')}}\, dx = \int_{\frac{\hat{c}}{\sqrt{w(\vec{a}')}}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx$$

$> \int_{\sqrt{3}}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx$. Hence $w(\vec{a}') > \frac{1}{3}\,\hat{c}^2$. Similarly

$z(\vec{a}') > \frac{1}{3}\left[1-\hat{c}\right]^2$. By Corollary 1, $\vec{a}'$ is a strict local minimum of $\beta w(\vec{a}) + (1-\beta)\, z(\vec{a})$ for some $0 < \beta < 1$.

IV. Third Order Solution for the Optimal Linear Discriminant

Suppose $x_1, x_2, \ldots, x_d$ are uncorrelated real random variables under each hypothesis with

$$E_1(x_i) = 0$$

$$E_2(x_i) = 1$$

$$E_1(x_i^2) = \lambda_1^{\,i}$$

$$E_2((x_i - 1)^2) = \lambda_2^{\,i}$$

In practical situations this can be achieved by applying the appropriate affine transformation to the data.

Let $\mathcal{L} = \left\{ \sum_2^d a_i x_i + (1 - \sum_2^d a_i) x_1;\ a_i \text{ real} \right\}$. Finding $L \in \mathcal{L}$ of minimum error is equivalent (in the third order sense) to finding critical points of

$$\beta_1^{\,2} E_1 \Big(\sum_2^d a_i x_i + (1 - \sum_2^d a_i) x_1 \Big)^2 + \beta_2^{\,2} E_2 \Big(\sum_2^d a_i (x_i - 1) + (1 - \sum_2^d a_i)(x_1 - 1) \Big)^2$$

$$+ \beta_1^{\,3} E_1 \Big(\sum_2^d a_i x_i + (1 - \sum_2^d a_i) x_1 \Big)^3 + \beta_2^{\,3} E_2 \Big(\sum_2^d a_i (x_i - 1) + (1 - \sum_2^d a_i)(x_1 - 1) \Big)^3$$

for various values of $\beta_i^{\,j}$, $-1 \le \beta_i^{\,j} \le +1$. This objective function is a cubic in $d-1$ variables and possesses in general 2^{d-1} critical points. However -

<u>Lemma 3</u> Let f be a cubic in $d-1$ dimensions. Then f has at most one strict local minimum.

<u>Proof</u>: Suppose f has two strict local minima, $\vec{x}$ and $\vec{y}$. Then f restricted to the line $\{\alpha\, \vec{x} + (1-\alpha)\, \vec{y};\ -\infty < \alpha < +\infty\}$ has strict local minima at $\alpha = 0$ and $\alpha = 1$. But the restriction of f is a cubic in one dimension which has at most one strict local minimum.

Suppose E_α is strictly concave at the values of the four parameters corresponding to the $L \in \mathcal{L}$ of minimum error. Then, by the preceding lemma, we need to determine at most one point in the domain of the objective function for each choice of $\beta_i^{\,j}$. In general the method of steepest descent will not yield a strict local minimum of a cubic since the cubic approaches both $\pm\infty$

far from the origin. However, if β_1^2 and β_2^2 are both positive and the skewnesses of the minimum L are small compared to the variances of that L, then the method of steepest descent may be used by starting at $\vec{a}_0$ which minimizes

$$\beta_1^2 E_1 \left(\sum_2^d a_i x_i + \left(1 - \sum_2^d a_i\right) x_1\right)^2$$

$$+ \beta_2^2 E_2 \left(\sum_2^d a_i (x_i - 1) + \left(1 - \sum_2^d a_i\right)(x_1 - 1)\right)^2$$

This is given by

$$a_{oi} = \frac{(\lambda_1^1 + \lambda_2^1)(\beta_1^2 \lambda_1^i + \beta_2^2 \lambda_2^i)^{-1}}{1 + (\lambda_1^1 + \lambda_2^1) \sum_{\ell=2}^{d} (\beta_1^2 \lambda_1^\ell + \beta_2^2 \lambda_2^\ell)^{-1}}$$

In other words we start at a (potential) second order solution * and use steepest descent to find a (potential) third order solution.

If $E_1 (x_i x_j x_k) = E_2 (x_i - 1)(x_j - 1)(x_k - 1) = 0$ except when $i = j = k$, the strict local minima of the above cubics may be explicitly determined under very mild restrictions. Let us assume throughout that $\beta_1^2 > 0$ and $\beta_2^2 > 0$. Our objective function becomes:

$$H(\vec{a}) = \sum_2^d a_i^2 (\beta_1^2 \lambda_1^i + \beta_2^2 \lambda_2^i) + \left(1 - \sum_2^d a_i\right)^2 (\beta_1^2 \lambda_1^1 + \beta_2^2 \lambda_2^1) +$$

*See Footnote (2) on page 18.

$$+ \sum_2^d a_i^3(\beta_1^3 E_1(x_i^3) + \beta_2^3 E_2(x_i-1)^3) + (1-\sum_2^d a_i)^3(\beta_1^3 E_1(x_1^3) + \beta_2^3 E_2(x_1-1)^3)$$

$$= \sum_2^d A_i a_i^2 + \sum_2^d B_i a_i^3 + A_1 (1-\sum_2^d a_i)^2 + B_1 (1-\sum_2^d a_i)^3$$

A strict local minimum of H, $\vec{a}$, corresponds to a critical point $(\vec{c},\Phi)$ of

$$K(\vec{c},\Phi) = \sum_1^d A_i c_i^2 + \sum_1^d B_i c_i^3 - \Phi(\sum_1^d c_i - 1)$$

where Φ is a Lagrange multiplier and

$$1 - \sum_2^d a_i = c_1,\ a_2 = c_2,\ \ldots,\ a_d = c_d.$$

Recall $A_i > 0$ since $\beta_1^2 > 0$, $\beta_2^2 > 0$. Differentiating K wrt c_i and setting the result equal to zero yields

$$2A_i c_i + 3B_i c_i^2 = \Phi$$

We attempt to solve the above system for $0 < \Phi < \Phi_{max}$ subject to the constraint $\sum_1^d c_i = 1$ where

$$\Phi_{max} = \min_{i;B_i<0} \frac{A_i^2}{3|B_i|} = \frac{A_{\tilde{k}}^2}{3|B_{\tilde{k}}|}^{*},$$

*See Footnote (3) on page 18.

taking c_i to be the smallest positive root. For a given Φ,

$$c_i = \begin{cases} \dfrac{-2A_i + \sqrt{4A_i^2 + 12\,B_i\Phi}}{6B_i} & , \quad B_i \neq 0 \\[2ex] \dfrac{\Phi}{2A_i} & , \quad B_i = 0 \quad . \end{cases}$$

Theorem 3 The system $2A_i\, c_i + 3B_i\, c_i^2 = \Phi$ has a solution of smallest positive roots for positive Φ with the roots summing to one provided

$$\sum_1^d \frac{-\,2A_i + \sqrt{4A_i^2 + 4B_i\, A_{\tilde{k}}^2\, (|B_{\tilde{k}}|)^{-1}}}{6B_i} > 1 \quad .^{*}$$

In addition the corresponding $\vec{a}$ is a strict local minimum of H.

Proof: For Φ close to zero the sum of the roots will be less than one. For $\Phi = \Phi_{max}$ the sum of the roots is greater than 1 by the condition of the theorem. By the mean value theorem there is Φ such that the corresponding roots sum to one.

Since $(\vec{c}, \Phi)$ is a critical point of K, the corresponding $\vec{a}$ critical point of H. Hence ∇H is zero at $\vec{a}$. To show that $\vec{a}$ is a

*See Footnote (4) on page 18.

strict local minimum we compute the Jacobian of H at $\vec{a}$ and show that it is positive definite:

$$\frac{\partial^2 H}{\partial a_i^{\,2}} = 2A_i + 6B_i\, a_i + 2A_1 + 6B_1 (1 - \sum_{2}^{d} a_i)$$

$$= \sqrt{4A_i^{\,2} + 12\, B_i \Phi} + \sqrt{4A_1 + 12\, B_1 \Phi}$$

$$\frac{\partial^2 H}{\partial a_i\, \partial a_j} = 2A_1 + 6B_1 (1 - \sum_{2}^{d} a_i) = \sqrt{4A_1^{\,2} + 12\, B_1 \Phi} \quad \text{for } i \neq j.$$

Since $\Phi < \Phi_{max}$ all the radical terms are positive. Hence $J(H) = \Lambda + \Omega$ where Λ is a diagonal matrix with positive eigenvalues and Ω is a matrix whose entries are a positive constant. Clearly such a matrix is positive definite. This completes the proof.

Footnotes

(1) In some cases one uses as a measure of error the probability of misclassification of one type given the probability of misclassification of the other. For such an error function the results of II remain valid. The results of III and IV in this setting will be discussed in a future paper.

(2) For $\beta_1^2 = \beta_2^2 = \frac{1}{2}$, this second order solution is known as the Fisher line.

(3) If there are negative B_i's, Φ_{max} is the largest Φ for which each quadratic has a solution. If there are no negative B_i's, we set $\Phi_{max} = +\infty$. Clearly each quadratic has a solution for $0 < \Phi < \Phi_{max}$.

(4) If $B_i = 0$ the i'th term is replaced by $\frac{\Phi_{max}}{2A_i}$. If all B_i are non-negative the theorem holds without the inequality.

References

(1) L. K. Jones, "On Optimal Discriminants between Two Classes of Random Variables in Terms of the Moments of their Distribution," submitted to SIAM Journal of Applied Mathematics.

(2) L. K. Jones, "Asymptotically Optimal Detector of Memory k for k-Dependent Random Signals," to appear in IEEE Trans. Inf. Theory.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ESD-TR-79-305 ✓	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) k^{th} Order Solutions to the Problem of Finding Optimal Discriminant Functions A06 0685 A070 455		5. TYPE OF REPORT & PERIOD COVERED Technical Note 6. PERFORMING ORG. REPORT NUMBER Technical Note 1980-4 ✓
7. AUTHOR(s) Lee K. Jones		8. CONTRACT OR GRANT NUMBER(s) F19628-80-C-0002 ✓
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. ✓ P.O. Box 73 Lexington, MA 02173		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 63311F Project No. 627A
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Systems Command, USAF Andrews AFB Washington, DC 20331		12. REPORT DATE 10 January 1980 13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

discriminant function
linear discriminant
normal distribution
concave error function

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This article describes a general method for discriminating between two random variables when we are required to use a discriminant function belonging to a given class.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)